Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluation
DOI: 10.54254/2755-2721/2025.PO24941

A Review on Deep Learning Applications in Medical Image Analysis

Xinyu Chen

Institute of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing, China
2023211387@bupt.cn

Abstract. In recent years, deep learning has demonstrated revolutionary impacts in medical image analysis due to its powerful feature learning and pattern recognition capabilities. This paper systematically reviews advances in core tasks and methodologies for medical image segmentation and classification. For segmentation tasks, this paper discusses how Fully Convolutional Networks (FCN) laid the foundation for pixel-level prediction, while their variants, such as the DeepLab series, optimized lesion segmentation accuracy through atrous convolution and multi-scale feature fusion. This research also covers U-Net and its 3D extensions (3D U-Net, V-Net), which significantly improved boundary consistency in organ segmentation by integrating skip connections and residual learning. Furthermore, this review examines Generative Adversarial Networks (SegAN, SCAN) and how they effectively addressed data scarcity and class imbalance through adversarial training. For classification tasks, this paper highlights classical convolutional neural networks like AlexNet, VGGNet, and ResNet, which achieved a paradigm shift from manual feature engineering to end-to-end learning via hierarchical feature abstraction and residual connections. Combined with transfer learning and multi-scale pooling strategies, these models substantially enhanced disease diagnosis accuracy and generalization. This review concludes that these technological breakthroughs have driven the transition of medical image analysis from traditional manual interpretation to intelligent, precise solutions, providing efficient and reliable support for clinical decision-making.

Keywords: Deep Learning, Medical Image Segmentation, Medical Image Classification

1. Introduction

Medical image analysis is a cornerstone of disease diagnosis and treatment planning, yet traditional methods relying on manual expertise suffer from inefficiency and inconsistency. The past decade has witnessed a transformative shift with the rise of deep learning, offering automated solutions for medical image analysis. Current research has made significant strides in areas such as organ segmentation [1-6] and disease classification [2,3,6,7], leveraging multi-layered neural networks to autonomously extract intricate image features. However, existing reviews often focus on individual techniques or specific applications, leaving a gap in understanding the overarching development

DOI: 10.54254/2755-2721/2025.PO24941

trajectory, critical technological iterations, and emerging future research directions within this rapidly evolving field.

This paper aims to bridge this gap by providing a comprehensive review of the fundamental advancements and applications of deep learning in medical image segmentation and classification. This research particularly seeks to identify the key milestones in deep learning's evolution for medical image analysis and project future trends that will shape its clinical impact. This study employs a systematic review methodology, analyzing seminal works and recent breakthroughs in deep learning architectures, focusing on their principles, evolutionary paths, and application outcomes in segmentation and classification tasks. By meticulously summarizing these developments, this paper intends to offer a robust theoretical framework for researchers and clinicians, fostering further innovation and accelerating the clinical translation of intelligent medical image analysis solutions. The insights presented here are crucial for understanding the present capabilities and future potential of deep learning in enhancing diagnostic precision and efficiency in healthcare.

2. Applications of deep learning in medical image analysis

In recent years, deep learning technology has demonstrated revolutionary impacts in the field of medical image analysis due to its powerful feature learning and pattern recognition capabilities. Its core roles can be summarized into the following directions: lesion and organ segmentation [1-8], image classification [2,3,6,7], object detection [3,6,7,8], and cross-modal image registration [3,4,8,9]. Furthermore, breakthroughs have been achieved in tasks such as image generation and enhancement [1,3-9] and disease prediction and diagnosis [2,3,7,9]. These advancements collectively facilitate the transition of medical imaging from traditional manual analysis to intelligent and precise analysis, providing efficient and reliable decision-making support for clinical practice.

2.1. The core tasks and methods in medical image analysis

2.1.1. Medical image segmentation

Medical image segmentation aims to precisely annotate target regions, providing pixel-level localization for diagnostic purposes. The mainstream methodologies and their evolution are summarized as follows:

Firstly, Fully Convolutional Networks (FCN) laid the foundation for end-to-end segmentation: by directly implementing pixel-level classification through fully convolutional layers, FCN has been widely applied in brain tumor and cardiac segmentation [1,4].Building upon this, the DeepLab series introduced dilated convolutions and Atrous Spatial Pyramid Pooling (ASPP), significantly enhancing the ability to segment multiscale lesions (e.g., microtumors) [1,3] while SegNet optimized edge precision of organs through its symmetric encoder-decoder architecture and pooling indices [1,2].

In parallel, U-Net and its variants have continuously innovated to address medical data characteristics: For instance, 3D U-Net extended convolutional operations to three dimensions, achieving a Dice coefficient >0.93 in hippocampal segmentation for Alzheimer’s disease [2,4] Notably, Attention U-Net integrated attention mechanisms to dynamically focus on critical regions (e.g., the pancreas), effectively improving sensitivity for small target segmentation [1,7]

DOI: 10.54254/2755-2721/2025.PO24941

Additionally, V-Net combined residual connections with Dice loss functions to refine boundary consistency in 3D liver segmentation [4,6]

On the other hand, Generative Adversarial Networks (GAN) offered novel solutions for data scarcity and class imbalance: SegAN employed multiscale L1 loss to markedly improve class balance in brain tumor segmentation [1,5].

2.1.2. Medical image classification

Medical image classification employs automated models to identify lesion types or anatomical structures.

Medical image classification has achieved a paradigm shift from manual feature design to automated learning through deep learning technologies. Classical CNN architectures have played a pivotal role in this domain: Firstly, AlexNet, as the first breakthrough model in the ImageNet competition, optimized the training process via ReLU activation functions and Dropout strategies, significantly improving the efficiency of diabetic retinopathy grading [2,3] Secondly, VGGNet enhanced feature extraction capabilities by stacking consecutive 3×3 convolutional layers and achieved high-precision classification in small-sample scenarios (e.g., breast nodule ultrasound detection) when combined with transfer learning strategies [2,7] Furthermore, ResNet effectively addressed the gradient degradation issue in deep networks through residual connections, attaining an AUC value exceeding 0.90 in benign-malignant pulmonary nodule screening tasks [3,6]. The combined use of these technologies not only increases classification accuracy but also speeds up the use of deep learning models in clinical settings.Technical improvements have further advanced the performance and practicality of classification models: On the one hand, Global Average Pooling (NIN) substantially reduced parameter counts by replacing fully connected layers, effectively mitigating overfitting [2,6] on the other hand, Spatial Pyramid Pooling (SPP) integrated multi-scale features, markedly enhancing model adaptability to medical image size variations [3,7]. Notably, transfer learning techniques successfully alleviated medical data scarcity by reusing generic features from ImageNet pre-trained models[2,3,7]The synergistic application of these technologies not only elevates classification accuracy but also accelerates the deployment efficiency of deep learning models in clinical practice.

3. Medical image segmentation based on deep learining

3.1. Fully convolutional neural networks [10]

3.1.1. FCN

CNN can only identify the category of the entire image, so it is not fit for segmention works. J. Long, E. Shelhamer and T. Darrell [10] use FCN to solve this problem.FCN uses three convolution layers which sizes are 7×7,1×1,1×1 to replace CNN's layer 5 to 7. Next, a softmax layer is behind to get the classfication of each pixel. Finally, CNN utilizes deconvolution layers to upsample the final convolutional feature maps to input image resolution, enabling pixel-wise prediction with preserved spatial information. The segmentation is achieved through pixel-level classification on reconstructed features. Therefore, the network can satisfy segmentation work.

DOI: 10.54254/2755-2721/2025.PO24941

3.1.2. Variants of FCN

The DeepLab series continuously optimized the shortcomings of FCN: DeepLabv1 [11] expanded the receptive field through atrous convolution while reducing pooling layers, combined with a conditional random field (CRF) to enhance boundary precision. DeepLabv2 [12] introduced the atrous spatial pyramid pooling (ASPP) module to capture multi-scale context and adopted ResNet-101 to strengthen feature representation. DeepLabv3 [13] further improved multi-scale object segmentation with cascaded/parallel atrous convolution modules and gradually phased out CRF. DeepLabv3+ [14] added a lightweight decoder to refine boundary details and integrated the Xception model with depthwise separable convolution for efficiency gains. Meanwhile, SegNet [15] constructed a symmetric encoder-decoder architecture, achieving non-learnable upsampling via pooling indices to preserve high-frequency features and reduce parameters, though its pooling restoration process risked losing local correlations.

3.2. U-Net

3.2.1. 2D and 3D U-Net

U-Net [16], introduced by Ronneberger et al. in 2015, is an encoder-decoder architecture derived from FCN and specifically designed for medical image segmentation. Its defining feature is a symmetric U-shaped structure with cross-layer skip connections. The encoder comprises four hierarchical submodules, each containing two convolutional layers followed by max-pooling to progressively downsample and extract features. The decoder reconstructs spatial resolution through upsampling while integrating skip-connected encoder features of corresponding resolutions, merging low-level details for precise localization and high-level semantics for robust feature abstraction. This design allows for detailed predictions at the pixel level, taking in images that are 572×572 pixels and producing outputs that are 388×388 pixels, which are fine-tuned for accuracy in medical segmentation tasks. Extending this framework to three-dimensional medical imaging, 3D U-Net [17] adopts volumetric convolutions and deconvolutions. The encoder utilizes 3×3×3 convolutions combined with pooling, while the decoder applies 2×2×2 deconvolutions followed by 3×3×3 convolutions to restore spatial context. It is made to take in 132×132×116 voxel inputs and produce 44×44×28 voxel outputs, keeping the details consistent across connected 2D slices and allowing training on datasets with few labels. Both U-Net variants eliminate fully connected layers, prioritizing lightweight computational efficiency without compromising segmentation precision, thereby establishing themselves as foundational models in medical image analysis.

3.2.2. V-Net

V-Net [18] is a 3D extension of U-Net designed for volumetric medical data segmentation. It introduces three key innovations: replacing traditional cross-entropy with the Dice coefficient loss function to address class imbalance, integrating residual learning by hierarchically summing input and output features across stages to enhance gradient propagation, and adopting a symmetric compression-decompression architecture. The compression path reduces resolution through stride adjustments to extract multi-scale 3D features using 5×5×5 convolutional kernels, while the decompression path restores spatial dimensions via deconvolution, merging low-resolution features to produce dual-channel segmentation outputs matching the original input size. By leveraging 1×1×1 convolutions for channel reduction, the model reduces computational overhead while preserving

DOI: 10.54254/2755-2721/2025.PO24941

voxel-level spatial coherence, achieving superior accuracy in 3D organ segmentation tasks such as prostate imaging.

3.3. Generative adversarial network

3.3.1. Segmentation adversarial network

SegAN [19] integrates U-Net as the generator within a generative adversarial network framework to address class imbalance in medical image segmentation. The architecture comprises two components: a segmentor network S based on U-Net and a critic network C, which are alternately trained through adversarial optimization. The segmentor utilizes a multi-scale L1 loss function, designed to outperform traditional single-scale losses, paired with a downsampling path using 4×4 convolutions at stride 2 and an upsampling path combining image resizing with a scaling factor of 2 followed by 3×3 convolutions at stride 1. The critic network evaluates ground truth-masked images against predictions from the segmentor, incentivizing boundary refinement. Tested on the BRATS brain tumor dataset, SegAN demonstrates improved robustness and segmentation accuracy, particularly in complex lesion regions, by leveraging adversarial training to address pixel-level class imbalance while maintaining structural coherence.

3.3.2. Structure correction adversarial network

SCAN [20] addresses lung field and heart segmentation in chest X-ray imaging, known as CXR, through an adversarial learning framework built on fully convolutional networks. The architecture consists of a segmentation network and a discriminator, both implemented as fully convolutional networks, eliminating dependency on pre-trained models like VGG and enabling training with only 247 annotated images. The segmentation network processes grayscale CXR directly, employing streamlined downsampling modules optimized for radiographic features. The discriminator incorporates physiological structural priors, using adversarial training to distinguish ground truth annotations from generated masks, thereby enforcing anatomically consistent segmentation. By combining detailed structural rules with a small amount of training data, SCAN improves the accuracy of anatomical features and provides an effective way to analyse medical images in places with limited resources.

4. Medical image classification based on deep learining

4.1. AlexNet

AlexNet [21] marked a critical turning point in deep learning by dominating the 2012 ImageNet Challenge with a novel eight-layer convolutional architecture. It has five layers that do convolution (using 11×11 and 5×5 filters) and three layers that are fully connected, and it brought in ReLU activation to speed up training, max-pooling to reduce size, and dropout (at a rate of 0.5) to prevent overfitting. Processing 227×227 RGB inputs, AlexNet establishes GPU acceleration and end-to-end feature learning as standards. This breakthrough directly catalyzed advancements like VGG and ResNet, cementing CNNs as the cornerstone of computer vision.

DOI: 10.54254/2755-2721/2025.PO24941

4.2. VGGNet

Building upon AlexNet's foundation, VGGNet [22] demonstrated the power of depth through uniform 3×3 convolutional stacks in its 16-/19-layer variants (VGG16/VGG19). By preserving spatial resolution via stride-1 convolutions and padding, each block combined multiple 3×3 layers with ReLU and 2×2 max-pooling, culminating in three fully connected layers. VGGNet proves that deeper networks with smaller kernels capture richer hierarchical features. Its modular design standardized deep layer stacking, enabling seamless transfer learning for tasks like object detection. The architecture's simplicity and depth-oriented philosophy directly inspired subsequent models, including ResNet and Inception.

4.3. ResNet

ResNet [23] addressed vanishing gradients in ultra-deep networks through residual blocks with skip connections, enabling stable training of 152-layer models (ResNet-152). Each block integrates stacked 3×3 convolutions, batch normalization, and ReLU, while identity shortcuts bypass nonlinear layers to propagate gradients. This innovation reduced ImageNet top-5 error to 3.57%, demonstrating that depth—when paired with residual learning—eliminates performance degradation. ResNet's paradigm shift decoupled depth from optimization barriers, inspiring derivatives like Wide ResNet and DenseNet. Its residual framework became ubiquitous across vision tasks, from detection to generation, proving that skip connections and extreme depth are pivotal for scaling modern AI systems.

5. Conclusion

Deep learning has profoundly transformed medical image analysis, significantly enhancing both accuracy and efficiency through continuous architectural innovations. This paper systematically reviewed key advancements in core tasks: for segmentation, FCNs laid the groundwork for precise pixel-level prediction, while U-Net and its 3D extensions pioneered robust solutions for intricate organ and lesion localization by incorporating skip connections. Furthermore, Generative Adversarial Networks (GANs) have become important tools that help solve problems of not having enough data and uneven class distribution by using a training method that involves competition, which is essential for delicate medical uses. For classification, the evolution of optimized CNN architectures like AlexNet, VGGNet, and ResNet, combined with transfer learning, revolutionized disease diagnosis by shifting from manual feature engineering to powerful end-to-end learning, leading to substantial improvements in diagnostic accuracy and generalization.

This review highlights that the evolution of deep learning in medical imaging is marked by critical iterative advancements, each addressing previous limitations and opening new possibilities. These important developments include FCN's change to fully convolutional layers, U-Net's clever use of skip connections to keep context, and GANs' creative methods for improving data and adapting to different areas. In classification, AlexNet's success validated GPU acceleration and deep architectures, VGGNet underscored the power of uniform depth, and ResNet critically solved the vanishing gradient problem, allowing for unprecedented model complexity and robustness. These advancements collectively constitute a vital development trajectory towards more sophisticated and reliable medical image analysis.

Looking ahead, the potential for deep learning in personalized medicine and intelligent diagnostics remains immense. Future research directions will focus on critical areas such as

DOI: 10.54254/2755-2721/2025.PO24941

developing more lightweight deployment strategies for efficient integration into clinical workflows and mobile health devices. Enhancing multi-modal data fusion, combining images with diverse patient information (e.g., genomics, EHRs), promises a more holistic understanding of disease. Importantly, making deep learning models easier to understand will help doctors trust them more and encourage their use in everyday medical practice, speeding up the move to smarter healthcare solutions.

References

[1] Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability, 13(3), 1224. https: //doi.org/10.3390/su13031224

[2] Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. Annals of translational medicine, 8(11), 713. https: //doi.org/10.21037/atm.2020.02.44

[3] Ker, J., Wang, L., Rao, J., & Lim, T. (2018). Deep learning applications in medical image analysis. IEEE Access, 6, 9375–9389. https: //doi.org/10.1109/access.2017.2788044

[4] Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19(1), 221–248. https: //doi.org/10.1146/annurev-bioeng-071516-044442

[5] Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Medical Image Analysis, 65, 101759. https: //doi.org/10.1016/j.media.2020.101759

[6] Suzuki, K. (2017). Overview of deep learning in medical imaging. Radiological Physics and Technology, 10(3), 257–273. https: //doi.org/10.1007/s12194-017-0406-5

[7] Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19(1), 221–248. https: //doi.org/10.1146/annurev-bioeng-071516-044442

[8] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on Deep Learning in medical image analysis. Medical Image Analysis, 42, 60–88. https: //doi.org/10.1016/j.media.2017.07.005

[9] Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift Für Medizinische Physik, 29(2), 102–127. https: //doi.org/10.1016/j.zemedi.2018.11.002

[10] Long, J., Shelhamer, E., & Darrell, T. (2015, June). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https: //doi.org/10.1109/cvpr.2015.7298965

[11] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv: 1412.7062. https: //doi.org/10.48550/arXiv.1412.7062

[12] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018b). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. https: //doi.org/10.1109/tpami.2017.2699184

[13] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv: 1706.05587. https: //doi.org/10.48550/arXiv.1706.05587

[14] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with atrous separable convolution for semantic image segmentation. In Lecture Notes in Computer Science (pp. 833–851). Springer International Publishing. https: //doi.org/10.1007/978-3-030-01234-2_49

[15] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. https: //doi.org/10.1109/tpami.2016.2644615

[16] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Lecture Notes in Computer Science (pp. 234–241). Springer International Publishing. https: //doi.org/10.1007/978-3-319-24574-4_28

[17] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D u-net: Learning dense volumetric segmentation from sparse annotation. In Lecture Notes in Computer Science (pp. 424–432). Springer International Publishing. https: //doi.org/10.1007/978-3-319-46723-8_49 October 17-21, 2016, Proceedings, Part II 19 (pp. 424-432). Springer International Publishing.

DOI: 10.54254/2755-2721/2025.PO24941

[18] Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV), 565–571. https: //doi.org/10.1109/3dv.2016.79
[19] Xue, Y., Xu, T., Zhang, H., Long, L. R., & Huang, X. (2018). SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation. Neuroinformatics, 16(3–4), 383–392. https: //doi.org/10.1007/s12021-018-9377-x
[20] Dai, W., Dong, N., Wang, Z., Liang, X., Zhang, H., & Xing, E. P. (2018). SCAN: Structure correcting adversarial network for organ segmentation in chest x-rays. In Lecture Notes in Computer Science (pp. 263–273). Springer International Publishing. https: //doi.org/10.1007/978-3-030-00889-5_30
[21] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https: //doi.org/10.1145/3065386
[22] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556. https: //doi.org/10.48550/arXiv.1409.1556
[23] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https: //doi.org/10.1109/cvpr.2016.90